QUEEN'S UNIVERSITY BELFAST

Dynamic Runtime Opcode Dataset Generation for Improving Malware Classification

Carlin, D., O'Kane, P., & Sezer, S. (2016). *Dynamic Runtime Opcode Dataset Generation for Improving Malware Classification*. Poster session presented at 19th International Symposium on Research in Attacks, Intrusions and Defenses, Evry, France.

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Publisher rights
© 2016 The Authors.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Download date:05. May. 2023

Cloud Malware

Domhnall Carlin, Dr. P. O'Kane & Prof. S. Sezer

Centre for Secure Information Technologies, Queen's University, Belfast, N. Ireland.

Research Objective

Develop a strategy for the detection of malware, which is immune to modern obfuscation methods, and which is applicable at the hypervisor level.

Motivation

Signature-detection is the most widely used approach in commercial malware detection. New malware must be captured & analysed for a signature, which is deployed to users. Obfuscation techniques compound the issue- 1000s of variations generated per day.

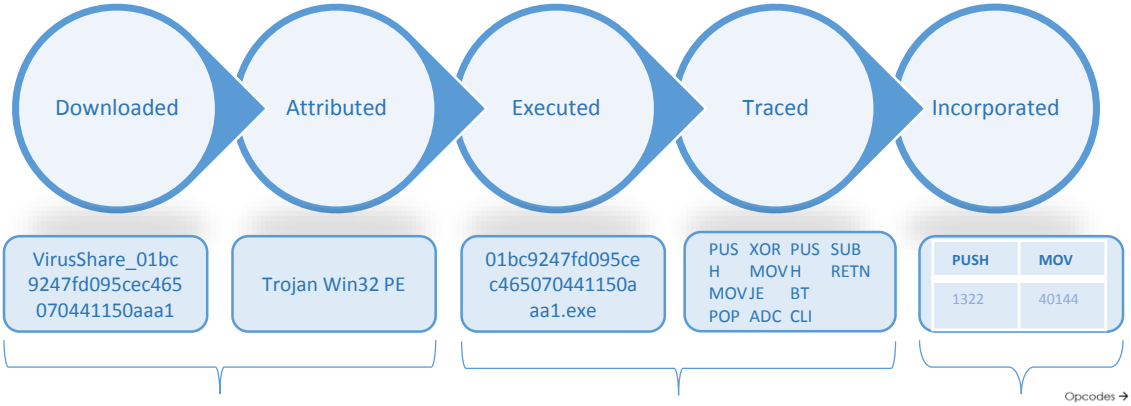
By definition, we are constantly behind the curve in the arms race against malware.

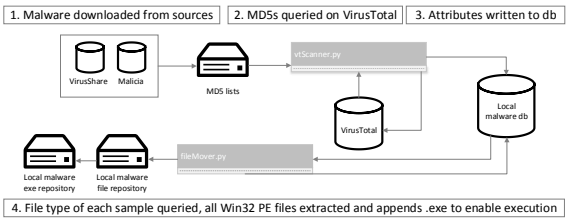
Relevant Research

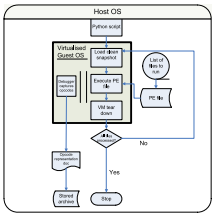
- Analysis of opcodes (assembly instructions) can detect malware, bypassing issues inherent in signature-based methods (i.e. the instance can be novel).
- Packed or encrypted malware cannot generally be investigated statically. Dynamic analysis allows the obfuscated malware to reveal itself.
- Datasets in the literature have been small, poorly sampled and prone to class imbalances.

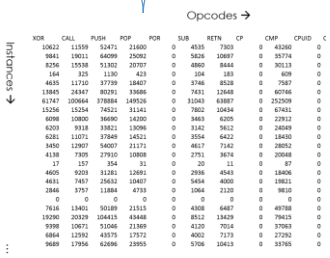
Creation of a large dynamically-generated run-trace dataset

Processing each malicious sample









Future work facilitated by new dataset:

Dataset stands at 48,252 executed, attributed and traced malware samples.

Unsupervised Deep Learning
Neural Net Classification
Relabelling based on opcode feature clusters

Optimal dataset composition
Algorithm comparisons
Novel feature selection &reduction

One-class learning
Rule-based classifiers
Dynamically uninvestigated malware types